DOCUMENT RESUME

ED 174 963 CS 004 956

AUTHOR Gregory, Russell D.
TITLE The Byte-Size Revolution: Computers and Readability.
PUB DATE Apr 79
NOTE 15p.; Paper presented at the Annual Meeting of the Western College Reading Association (12th, Honolulu, Hawaii, April 7-10, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Community Colleges; *Computer Programs; Instructional Materials; Junior Colleges; *Readability Formulas; *Reading Programs; *Word Lists

ABSTRACT

This report describes the design and operation of a computer readability service at Linn-Benton Community College in Albany, Oregon, which has evolved from a program containing a single readability formula to its present six-formula capacity. The daily operation of the service and the results provided by the computer are described in detail, and it is noted that in addition to several grade level estimates for the materials analyzed, the program developed can also identify words at or above a predetermined grade level on a list of 9,000 words. A comparison of human versus machine analysis is given. The report concludes that a computer readability service is a practical campus service that could be equally as important as other developmental services. (Author/MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED174963

THE BYTE*-SIZE REVOLUTION: COMPUTERS AND READABILITY

Russell D. Gregory
Developmental Center Instructor

Linn-Benton Community College
Albany, Oregon

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Russell D. Gregory

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

*A computer's main memory is divided into a number of memory locations. Each location can store a certain number of bits. This number varies from one computer to another, although eight bits, called a byte, is common on many modern machines.

CS004956

ABSTRACT

This report describes the design and operation of a computer readability service at Linn-Benton Community College in Albany, Oregon. The service has evolved from a program containing a single formula to its present six formula capacity. The daily operation of the service and the results provided by the computer are described in detail. A comparison of human versus machine analysis is given. The conclusion is that computer readability is a practical campus service that could be equally as important as other developmental services.

Purpose

The purpose of this paper is to describe the design and operation of a campus computer readability service. This report is written for those who already have some knowledge of readability formulas. It is especially appropriate for those who have applied the formulas by hand. The information will be most helpful to those who have access to a computer. While all of the information necessary to establish a computer readability system is not given, the report does help to explain the system as it now exists at Linn-Benton Community College (LBCC), in Albany, Oregon.

Background

The data in this report were obtained from research articles and from the experience gained in designing and operating a computer readability system. The system has existed in its present form since September, 1978, although earlier versions have been used at LBCC since April, 1976.

Although some recent articles have questioned the appropriateness of using readability formulas to analyze certain types of written materials, it is not the purpose of this paper to defend or support their use. It is wise, however, to at least be aware of some of the criticism that has been leveled at readability formulas. Nelson (24) warns of the formulas'

limitations when used by content-area instructors. Standal (26:546) writes, "They Readability formulas are not absolute. If they are regarded as general indicators, they can be quite useful." Others, such as Maxwell (22) and Marshall (20,21), cite sections of prose that defy accurate labeling by readability formulas.

After using our computer readability service to find the difficulty levels of hundreds of materials, we believe it is the use of the information that should concern readability analysts, not the method of application. If readability formulas can accurately be applied to large quantities of material by a computer in much less time than is required by a human analyst, then the analyst has much more time to spend on other tasks. The analyst can determine who the audience is, what the specialized terms are, and how the other statistics generated by the computer can be used to assist readers of the material.

A 1975 article by Barry and Stevenson (1) stirred this author's curiosity enough to start the preliminary steps toward the design of a computer readability system. In replicating Barry and Stevenson's work, we made some changes that made the program more useful to us. One of these changes was a decision to print out a list of words not found in the Dale-Chall list. We found the list to be helpful when working with students who were assigned paragraphs that had previously been analyzed by the computer. For several months, this single-formula program comprised the entire LBCC "system."

Further research revealed few references to previous computer readability studies. Much of the research was contained in unpublished dissertations. Studies that provided us with more information about determining readability by computer include those done by: Berkeley (2), Branch (4), Coke and Rothkopf (5), Danielson and Bryan (7), Fang (10), and General Motors (16). Klare (18) provided a list of readability studies that had been done prior to 1975.

An early goal was to expand the computer program to include other readability

formulas. The first major hurdle was the requirement by some formula authors such as Fry (14, 15), Flesch (13), McGlaughlin (23), and Farr, Jenkins, and Paterson (11) that syllables should be counted within each sample. The high costs of data processing in the 1960's and complex rules of English syllabication caused at least one researcher, Fang (10), to conclude that such a computer program was of questionable value. Nevertheless, Fang did design a syllable-counting computer program that contained many rules, exceptions to the rules, and exceptions to the exceptions. His program was designed to test broadcast news copy and included some rules that seemed inappropriate for readability formulas. According to his guidelines, for example, the word "fire" has two syllables. The next steps were to cull those rules that were inapplicable, apply the formula to hundreds of words from dozens of subject areas, and make adjustments to the rules.

Comparisons of computer syllable counts with human counts taken from the same passages have yielded correlations of .98. This is despite the wide range of subject matter analyzed, including monosyllabic grade school materials and polysyllabic technical program materials. Although other computer readability designers, such as General Motors (16) and Felsenthal and others (12) suggest using mathematical estimates to determine syllables based on letter and vowel counts, we have not found this method to be as accurate as ours. Once the syllable-counting subprogram was made as efficient as possible, those formulas Fry (14), Flesch (13), McGlaughlin (23), and Farr-Jenkins-Paterson (11) requiring this step were added to the system.

The last formula to be added to the system was one designed by Spache (25). This formula determines grade levels of difficulty for materials written below the fourth grade level. Since the Spache formula also requires a check of each word against a list of familiar words, we then consolidated the Spache, Dale-Chall, and Taylor (27) lists.

The consolidated word list can also be accessed for uses other than readability estimates. For example, a user can have the computer print out all the words at the eighth grade level (as identified by Taylor) that have a common letter combination such as "ie-ei." This data can help spelling instructors focus in on the words most likely to be used by students of varying abilities.

Operation Of The Readability Service

The LBCC Computer Readability Service consists of a one-fifth time reading specialist, a data entry operator, and the computer readability system. The specialist arranges for materials to be analyzed by working with faculty members or others who request the service. The specialist can also make arrangements with the bookstore to borrow new materials for a few hours.

Once the materials to be analyzed are in the possession of the reading specialist, two methods of handling them are possible. Excerpts may be taken directly from the materials or photocopies may be made of the materials. The latter technique is usually more desirable, since materials can then be returned to the requestor almost immediately. The guaranteed early return of materials is often the factor that causes people to use the service, since instructors, for instance, may need their textbooks almost nightly for class preparation.

Most formula authors refer to sample selection as being a "random" process. We use a statistical random process to choose samples from throughout the materials. When samples must be taken from specified sections or chapters, we use a stratified random sampling procedure. This ensures a fair representation of samples within each section or chapter. In any case, we try to avoid picking random samples based solely on personal preference. If, however, a requestor wants to know the level of difficulty for a specific section of prose, we also will analyze selected samples by computer.

After the sample pages have been picked and copies have been made of them, the data entry person enters the samples into the computer via terminal or key punched computer cards. Following a few simple rules for input similar to those proposed by Barry and Stevenson (1), the entry person can accomplish this step in no more time than it would take to type the material. One advantage to using cards is that a permanent record can be kept for later, more detailed analysis.

The computer time necessary for the readability analysis varies in relation to the amount of information that has been requested. For instance, if only the Fry and Flesch formulas are requested, less than ten seconds are required for analysis of an entire 15-sample selection. If formulas that require extensive word list checks are requested, however, the computer time increases. When the computer is instructed to print out the samples, provide the percentage of personal pronouns denoting gender, and list the words at or above a certain grade level, the processing time increases even more. Still, this maximum minute-per-sample time is impressive when contrasted with the several hours it would take a human analyst to count syllables, words, and sentences (possibly rechecking each several times), match words in lists, and perform mathematical computations.

The computer printouts received by the requestor are depicted in figures (I and II). Each sample summary provides readability information for just one passage. The Selection Summary is a composite of the sample summaries. The printouts have been designed to minimize the necessity for further analysis. Their 8½ X 11 inch size provides an easily copied and stored product. The various title elements allow the user to set up files by several different categories such as title, subject area, publisher, author, and grade level.

The passage lengths may vary slightly, but all statistics are based on a hypothetical 100-word sample. Sample sizes can be as large as 200 words. There is no minimum sample size, although passages with less than 75 words

FIGURE I
READABILITY ANALYSIS
RUSS GREGORY
DEVELOPMENT OFFICE
LBCC

****
* SELECTION *
* SUMMARY *
****

TITLE: HOW TO TAKE TESTS
AUTHOR: WALTER PAUK
PUBLISHER: MCGRAW HILL
SUBJECT AREA: STUDY SKILLS

SELECTION CHARACTERISTICS

SENTENCE LENGTH
FREQUENCY
1-2 (WORDS)..... 0%
3-5..... 11
6-12..... 28
13-19..... 37
20+..... 24
Mean (WORDS).....17

WORD LENGTH
FREQUENCY
1-3 (LETTERS)..... 42%
4-5..... 27
6-8..... 22
9-11..... 8
12+.....
1 (SYLLABLES)..... 66%
2..... 21
3..... 8
4..... 3
5+..... 1
MEAN (LETTERS).....5
MEAN (SYLLABLES).....1

GENDER OF SELECTED PERSONAL PRONOUNS
MALE..... 100%
FEMALE..... 0

UNFAMILIAR WORDS
DALE-CHALL LIST..... 18%
SPACHE LIST..... 22

WORD GRADE LEVEL

UN- 4%	3 - 8%	7 - 2%	11 - 0%
PP- 32	4 - 5	8 - 2	12 - 0
1 - 26	5 - 4	9 - 0	CL - 0
2 - 12	6 - 4	10 - 1	

FIGURE I
(Continued)

GRADE LEVEL READING EASE

FORMULA	GRADE	RANGE
FARR-JENKINS-PATERSON (5 THRU CG)	10.9	
FLESCH (5 THRU CG)	10.0	
SMOG (PP AND UP)	10.9	
FRY (1 THRU CG)	8.8	
DALE-CHALL (4-CG)	9.8	
MEAN (FOR SCORES 5 THRU CG INCLUSIVE)	10.1	2.0
PUBLISHERS DESIGNATED GRADE LEVEL	0	

FIGURE II
READABILITY ANALYSIS
RUSS GREGORY
DEVELOPMENT OFFICE
LBCC

TITLE: HOW TO TAKE TESTS
AUTHOR: WALTER PAUK
PUBLISHER: MCGRAW HILL
SUBJECT AREA: STUDY SKILLS
PASSAGE NUMBER: 10 (CHAPTER 0, PAGE 106)

PASSAGE

IN A WAY ALL THE OPTIONS HAVE A CERTAIN APPEAL. BY ATTENDING TO THEIR DIFFERENCES AND MORE EXACT MEANINGS YOU WILL BE BETTER ABLE TO CHOOSE THE CORRECT ANSWER. BLUNDER IMPLIES GROSS STUPIDITY OR IGNORANCE DECISION CONVEYS THAT A CONSCIOUS CHOICE WAS MADE FAULT REFERS TO AN IMPERFECTION OR BLAMEWORTHY DEFECT AND MISTAKE EXPRESSES A MISCONCEPTION OR MISUNDERSTANDING LIKE THE ONE IN THIS SAMPLE. OPTION D IS CORRECT.

READING CHARACTERISTICS

SENTENCES	4	
WORDS	68	
SENTENCE LENGTH		
FREQUENCY		
1-2 (WORDS)	0	0%
3-5	1	25
6-12	1	25
13-19	1	25
20+	1	25
MEAN (WORDS)	17	
WORD LENGTH		
FREQUENCY		
1-3 (LETTERS)	27	40%
4-5	12	18
6-8	20	28
9-11	6	9
12+	3	4
1 (SYLLABLES)	39	57%
2	19	28
3	5	7
4	4	6
5+	1	1
MEAN (LETTERS)	5	
MEAN (SYLLABLES)	2	
GENDER OF SELECTED PERSONAL PRONOUNS		
MALE	0	0%
FEMALE	0	0

FIGURE II
(Continued)

UNFAMILIAR WORDS
DALE-CHALL LIST.....21 31%
SPACHE LIST.....24 36

WORD GRADE LEVEL

UN-	5	7%	3 -	7	10%	7 -	5	7%	11 -	0	0%
PP-	20	30	4 -	1	1	8 -	1	1	12 -	0	0
1 -	15	22	5 -	6	9	9 -	1	1	CL -	0	0
2 -	2	3	6 -	3	4	10 -	1	1			

WORDS AT OR ABOVE GRADE LEVEL 8

MEANINGS (UN) IMPLIES (9) STUPIDITY (UN) IMPERFECTION (UN) BLAMEWORTHY (UN)
DEFECT (8) MISCONCEPTION (10) D (UN)

GRADE LEVEL READING EASE

FARR-JENKINS-PATERSON (5 THRU CG).....13.5
FLESCH (5 THRU CG).....12.3
SMOG (PP AND UP).....12.0
FRY (1 THRU CG).....13.0
DALE-CHALL (4-CG).....14.2

RANGE (FOR SCORES 5 THRU CG INCLUSIVE)..... 2.2
MEAN (FOR SCORES 5 THRU CG INCLUSIVE).....13.0

appear to be artificially high or low in difficulty when compared with other samples of greater size taken from the same materials.

One especially useful application of the computer system has been to provide readability analyses of learning activity packets. When these individualized instructional materials are analyzed prior to publication, the information can help provide the writers with suggestions for modifying the packets. For example, readability analyses of materials for a state-funded culinary arts curriculum were instrumental in causing the learning packet instructions to be rewritten.

One other feature of the computer system is the identification of words at or above a predetermined grade level based on Taylor's list. If the requestor is interested in finding out what words are at or above the eighth grade level, for instance, the computer will print out those words along with the words not found in Taylor's 9,000 word list. This list may help writers of teacher-made materials prepare the readers for difficult words. The changes made by writers may not appreciably affect the readability grade levels of the materials as measured by formulas, but the materials may be easier to understand because of the changes.

Conclusions

The feasibility of offering computer readability assessments as a campus developmental service has been demonstrated by the system at LBCC. By applying research findings to an area that has been traditionally offered, though usually not developed into a campus service, LBCC's computer readability system provides a model for other institutions.

One advantage to having a computer analyze materials for readability is that the readability process can be looked at more objectively. The human analyst becomes involved with the <u>results</u> of readability studies rather than the <u>process</u> of readability. The analyst can also avoid the misconception that

a goal has been reached when a list of grade levels has been handed over to someone who requests a readability analysis.

The old computer saw of garbage-in, garbage-out (GIGO) applies with equal emphasis to readability analyses. Materials submitted for computer analysis should be ready for publication before being entered. A recent example serves to point this out. The readability levels of some learning packets on the use of basic tools were hovering around the eighth grade mark. The formulas had been applied to the rough drafts of the materials, however, and further checks for evidence of lucid style and logical sequence had not yet been made. These checks revealed that some sections of the materials, though written at an "easy" level, were nearly incomprehensible. Used as a tool, the computer can provide valuable assistance to the reading specialist and writer. Used as a crutch, it can provide a wealth of misleading data.

BIBLIOGRAPHY

1. Barry, Jeanne Gardner & Stevenson, Timothy E., Using a computer to calculate the Dale-Chall Formula. Journal of Reading. December, 1975, 19, 218-222.

2. Berkeley, Edmund C., Computer-assisted Explanations. Cambridge: Information International, 1967.

3. Bond, G.L., and Tinker, M.A. Reading Difficulties: Their Diagnosis and Correction. New York: Meredith Corporation, 1973.

4. Branch, Lee Edward. The Dale-Chall Readability Formula As Programmed For The Computer. Unpublished doctoral dissertation. Claremont Graduate School, 1972.

5. Coke, Esther U. & Rothkopf, Ernst Z. Note on a simple algorithm for a computer-produced reading ease score. Journal of Applied Psychology, June 1970, 54, 208-210.

6. Dale, Edgar & Chall, Jeanne S. A formula for predicting readability. Educational Research Bulletin, January 21 and February 17, 1978, 27, 11-20, 37-54.

7. Danielson, Wayne A. & Bryan, Sam Dunn. Computer automation of two readability formulas. Journalism Quarterly, Spring 1963, 39, 201-206.

8. Eberwein, Lowell D. The variability of readability of basal reader textbooks and how much teachers know about it. Reading World, March 1979, 18, 259-272.

9. Entin, Eileen B. & Klare, George R. Factor analyses of three correlation matrices of readability variables. Journal of Reading Behavior, Fall 1978, 10, 279-290.

10. Fang, Irving E. By computer: Flesch's Reading Ease score and a syllable counter. Behavioral Science, May 1968, 13, 249-251.

11. Farr, James N., Jenkins, James J., & Paterson, Donald G. Simplification of Flesch Reading Ease Formula. Journal of Applied Psychology, October 1951, 35, 333-337.

12. Felsenthal, Norman, Shamo, Wayne G., & Bittner, John R. A comparison of award-winning radio commercials with their day-to-day counterparts. Journal of Broadcasting, 15, Summer 1971, 309-315.

13. Flesch, Rudolf. A new readability yardstick. Journal of Applied Psychology, 32, June 1948, 221-233.

14. Fry, Edward B. A readability formula that saves time. Journal of Reading. April 1968, 11, 513-516, 575-578.

15. Fry, Edward B. Fry's readability graph: clarifications, validity, and extension to level 17. Journal of Reading, 21, December 1977, 242-252.

16. General Motors (Marketing Staff, Service Research Group). STAR: General Motors computerized simple test approach for readability, a tool for improved communications. Undated.

17. Harris, Albert J. & Jacobson, Milton D. A framework for readability research: moving beyond Herbert Spencer. Journal of Reading, February 1979, 22, 390-398.

18. Klare, George. Assessing readability. Reading Research Quarterly, 1974-75, 10, 62-102.

19. Konshak, Dennis J. Readability's Usefulness for developmental composition. In Gwyn Enright (Ed.) Proceedings of the Eleventh Annual Conference of the Western College Reading Association, 1978, 11, 105-110.

20. Marshall, Nancy. Readability and comprehensibility. Journal of Reading. March 1979, 22, 542-544.

21. Marshall, Nancy & Glock, Marvin D. Comprehension of corrected discourse: a study into the relationships between the structure of text and information recalled. Reading Research Quarterly, 1978-79, 14, 10-56.

22. Maxwell, Martha. Readability: have we gone too far? Journal of Reading, March 1978, 21, 525-530.

23. McGlaughlin, G. Harry. SMOG grading--a new readability formula. Journal of Reading, May 1969, 12, 639-646.

24. Nelson, Joan. Readability: Some cautions for the content area teacher. Journal of Reading, April 1978, 21, 620-625.

25. Spache, George. A new readability formula for primary grade reading materials. Elementary School Journal, March 1953, 53, 410-413.

26. Standal, Timothy C. Readability formulas: what's out, what's in? The Reading Teacher, March 1978, 31, 642-646.

27. Taylor, Stanford E. & Frackenpohl, Helen & White, Catherine E. A revised core vocabulary: a basic vocabulary for grades 1-8, an advanced vocabulary for grades 9-13. Research and Information Bulletin No. 5 (revised). New York: Educational Development Laboratories, 1969.

28. Worthington, James S. The readability of footnotes to financial statements and how to improve them. Journal of Reading, 20, March 1977, 469-479.